Report 77-03
Stanford -- KSL

Scientific DataLink

Rule-Based Medical Decision Making by Computer. Bruce G. Buchanan, Randall Davis, Victor L. Yu, Stanley Cohen, 1977

card 1 of 1

Submitted to MEDINFO.77
HPP-77-3

Rule Based Medical Decision Making by Computer

Bruce G. Buchanan, Randall Davis, Victor Yu, and Stanley Cohen(1)
Stanford University
Stanford, California, U.S.A.

A new approach to medical decision making by computers is summarized. Individual inference rules -- both definitional and judgmental -- are incorporated in a symbolic reasoning program, named MYCIN, that is designed to provide assistance to physicians regarding appropriate therapy for infectious diseases. Because of the structure of the knowledge base, the program can explain its reasoning and can easily integrate new items of information about the domain. Results of a formal study show that the program provides good advice most of the time, as judged by ten experts in the field.

1 BACKGROUND

A number of recent studies indicate a major need to improve the quality of antimicrobial therapy. Almost one-half of the total cost of drugs spent in treating hospitalized patients is spend on antibiotics [1,2], and a significant part of this therapy is associated with serious misuse [2,3,4,5]. One problem involves incorrect selection of a therapeutic regimen [4], while another involves the incorrect decision to administer any antibiotic [2,4,5]. For example, one recent study concluded that one out of every four people in the United States was given penicillin during a recent year, and nearly 90% of these prescriptions were unnecessary [6].

---

(1) We gratefully acknowledge the assistance of many others who participated in the development of the MYCIN program, including J.Aikins, S.Axline, R.Blum, R.Chavez-Pardo, W.Clancey, L.Fagan, J.Hannigan, F.Rhame, C.Scott, E.H.Shortliffe, S.Wraith, and W.VanMelle. This work was supported in part by the Bureau of Health Services, Research and Evaluation (Grant No. HS-01544) and the Advanced Research Projects Agency (Order No. 2494).

2 THE MYCIN SYSTEM

Comprehensive reviews of our work appear in previous publications [11,12]. Briefly, we have developed a computer program named MYCIN that can offer consultative advice on the diagnosis and therapy selection for bacteremia and meningitis, two areas central to the management of infectious disease. (Knowledge of other infectious diseases will be added in the future.) This work has been guided by three fundamental objectives.

(1) A major goal of the MYCIN system has been to provide a computer-based therapeutic tool designed to be clinically useful. This requires development of a system that has a medically sound knowledge base, and that displays a high level of clinical competence in its field. The program must first convince clinicians of the quality of the information it is providing before they will be willing to use it.

(2) Since many clinicians will not accept advice from a computer-based system unless they can understand why the therapy has been selected, the system must do more than pontificate. It should have the ability to explain the reasoning behind its decisions, in terms that suggest to the physician that the program approaches the problem as he does. This permits the physician to validate the program's reasoning, and modify (or reject) the advice if the reasoning is not sound. It also gives the program an inherent instructional capability that allows the physician to learn from each consultation session.

(3) Third, we want the program to learn from clinical experts. The system therefore must be able to acquire knowledge by interacting with experts and incorporate it into the knowledge base.

Three inter-related parts of the MYCIN system accomplish these goals. The <u>consultation system</u> uses the knowledge base, along with patient-related data entered by the physician to generate therapeutic advice. The <u>explanation system</u> has the ability to explain the reasoning used during the consultation, and to document the motivation for questions asked or the rationale for conclusions reached. Finally, the <u>knowledge acquisition system</u> enables experts in antimicrobial therapy to update MYCIN's knowledge base, without requiring that they know how to program a computer.

A principal feature of MYCIN central to these goals is the format in which its knowledge is encoded. The MYCIN knowledge base currently contains approximately 400 such rules. Each rule consists of a set of preconditions (the "premises") which, if true, justify the conclusion made in the "action" part of the rule. These diagnostic and therapeutic decision rules have been formulated during extensive discussions of clinical case histories. One example is shown below.

Rule Based Decision Making: MYCIN

If 1) the gram stain of the organism is gram negative, and
2) the morphology of the organism is rod, and
3) the aerobicity of the organism is anaerobic,
then there is suggestive evidence (.6) that the identity of the organism is Bacteroides.

Many of the system's unique and important capabilities are made possible by encoding knowledge in comprehensive, modular "chunks" such as this.

Consultation. The consultation system uses its collection of rules to make conclusions about the patient. If, for, instance, it is attempting to determine the identity of an organism responsible for a particular infection, it retrieves the entire list of rules which, like the one above, conclude about identity. It then attempts to ascertain whether the conclusion of the first rule is valid, by evaluating in turn each of the premises. Thus, for the rule above, the first thing to find out is gram stain. If this information is already available in the data base, the program retrieves it. If not, determination of gram stain becomes the objective of a new rule, and the program retrieves all rules which conclude about it, and tries to use each of them to obtain the value of gram stain. If, after trying all the relevant rules, the answer still has not been discovered, the program asks the user for the relevant clinical information which will permit it to establish the validity of the premise clause. Thus, the rules "unwind" to produce a succession of goals, and it is the attempt to achieve each goal that drives the consultation.

Explanation. The use of a rule-based representation of knowledge makes it possible for the system to explain the basis for its clinical recommendations. For example, if the clinician asks "How did you determine the identity of the organism?" the program answers by displaying the rules which were actually used, and explaining, if requested, how each premise was established. This provides a far more comprehensible and acceptable explanation than would be possible if the program were using a simple statistical approach to diagnosis.

Knowledge Acquisition. An expert can give the program new "chunks" of information in much the same way he gives students knowledge about his field. This rule-based representation of knowledge means that the expert himself can make the program more competent, without having to know anything about computer programming. In addition, since the rules are independent of one another, and are used by the program whenever appropriate, the addition of a new rule (or modification of an existing rule) does not require altering other items in the knowledge base, as is often necessary with systems using the decision-tree methodology.

Other benefits gained from this approach have been explained in more detail in the references [11,12].

2.1 Other Approaches

There are three other approaches to the problem of diagnosis and therapy selection that have received extensive attention in the literature:

(i) Decision trees - as in [7], in which a sequence of decisions is structured in the form of a tree. Each node represents a particular question, whose answer determines which branch of the tree to follow to get to the next question. Final results are obtained by descending all the way to a leaf of the tree.

(ii) Bayesian techniques - as in [8], in which extensive frequency data make it possible to use Bayes' theorem as a basis for diagnosis.

(iii) Decision analysis and utility theory - as in [9], in which there is associated with each piece of information a likely cost of obtaining it, and a measure of the benefit to be derived from having it. Information is requested until the projected cost of asking another question (perhaps requiring another lab test or operative procedure) outweighs the benefit (in terms of a more precise diagnosis) to be obtained.

Each of these has a number of attractive aspects, but also encounters some limitations which provided the motivation for our investigation of a rule-based system. Decision trees, for example, offer simple, readily understandable procedures for diagnosing specific ailments. Problems occur, however, if they encounter unexpected data or if test results are unavailable. The representation of knowledge they offer can be somewhat inflexible, as well, since the attempt to make changes deep down in the tree often requires consideration of all previous decisions made further up the tree.

The Bayesian technique offers an appealing generality and precision, since it is a domain independent technique based on exact principles. Limitations here arise from the need for extensive amounts of frequency data concerning a priori and conditional probabilities. Where these data exist, the technique can be used quite effectively, but such figures may not often be available [10].

Techniques based on utility theory can present a well-motivated sequence of questions that appears to "zero in" on the underlying ailment. Like the Bayesian approach, however, it requires on extensive data on conditional probabilities of symptoms and disease.

Rule Based Decision Making: MYCIN

While all of these techniques present compact encodings of knowledge that can provide an appealing efficiency to programs based on them, there is an unavoidable loss of comprehensibility to the physician using them. Reasoning which requires several distinct inferential steps by a clinician, for instance, might be expressed in a single conditional probability statement in the Bayesian method.

Finally, we place great emphasis on the flexibility of the knowledge base. A substantial amount of knowledge is required to support a high level of performance, and this means that modification and augmentation of the knowledge base will continue for an extended period. Each modification must therefore be a reasonable task, or the program will soon begin to stagnate. A flexible knowledge base also means that the system is inherently dynamic in character. It is easily modified to take into account regional variations in practice, new results which arise from progress in medical research, and changes in drug resistance patterns.

3 EVALUATION OF MYCIN

To measure our progress on building a competent aid to physicians, we compared MYCIN's performance against the performance of experts from whom physicians ordinarily seek consultative advice. The same set of clinical data was provided to both MYCIN and ten experts in infectious disease therapy. [Five of the experts were nationally recognized authorities not at Stanford, the other five were clinical fellows in the Infectious Disease Division at Stanford.] The judgments of the program and the experts were compared on the questions of:

1) organism significance

2) organism identity

3) appropriate therapy

In addition, the experts were asked to evaluate MYCIN's performance with respect to

4) overall performance evaluation

5) potential impact as a clinical tool or teaching facility

3.1 Details of the Evaluation Procedure

Consecutive patients with positive blood samples were screened until we obtained at least 10 patients for whom MYCIN recommended therapy, and 15 patients overall (Patients were rejected if they were outpatients when the sample was drawn, if they had a previous blood culture in the preceding seven days, or if they had a diagnosis of meningitis or infectious endocarditis.) For each of the patients accepted, a one to two page clinical summary was prepared and combined with a summary of the laboratory test data as of the time when the first blood culture was obtained. This information was then used to obtain a therapeutic evaluation from MYCIN.

Each of the participating experts received a set of fifteen evaluation forms (one for each patient), including the clinical summary and lab data. Before looking at MYCIN's conclusions, each expert recorded his own conclusions about the infections, likely infecting organisms, and appropriate therapy based on the same information supplied to MYCIN. This allowed us to compare the expert's answers to MYCIN's. We also asked for the expert's opinion of the system's performance. Since it is difficult to select a single number which summarizes performance, we have in general measured each of the parameters listed above in three ways: (i) the percent of instances in which the program was judged exactly correct, (ii) the percent of instances in which the program's performance was judged exactly correct or an acceptable alternative, and (iii) the percent of cases in which a majority of the experts judged its performance exactly correct or an acceptable alternative. By using all three measures, we obtain a range of figures which give a good picture of the program's performance.

All of these attempts to evaluate performance are complicated by the fact that (as expected) the experts' own choices about each patient were not unanimous. Thus, we cannot ask whether MYCIN's answers were "correct" in any absolute sense, since there was no agreement on what constitutes "correct". Instead, we ask how often each individual expert rated the program's responses as correct. Without a second round of questionnaires, which would ask each expert to rate the acceptability of the other 9 experts' responses, we lack direct information about intra-expert variability. The figures below should be reviewed with this caveat in mind. Table I summarizes the first three measurements.

A. Organism Significance. The first question on the evaluation form gave the expert a chance to indicate whether or not he felt the patient needed to be treated. The first column of row one indicates the number of times the expert indicated no treatment was necessary for a case in which MYCIN also judged the organism to be a contaminant.

MYCIN decided four of 15 patients had organisms in their blood cultures which were not likely to be pathogens and therefore recommended no therapy. The majority of infectious diseases experts independently arrived at the same conclusion on these four patients. In only 7.5% (3/40) of instances was there disagreement among experts with MYCIN's decision on whether or not to treat. In the remaining eleven patients, MYCIN recommended therapy and the ten experts were unanimous in agreeing that all eleven of these patients required therapy.

B. <u>Organism</u> <u>Identity</u>. For organism identity, the experts were asked to rate each of MYCIN's selections as exactly correct (they agreed that the organism was likely to be present), an acceptable alternative (they had not chosen that organism, but agreed it might be present), or an unacceptable choice (they disagreed with its selection).

The second row of table 1 summarizes the data in this area. In the eleven treated cases, there was a total of 46 organisms that MYCIN selected as requiring therapy. (With ten experts rating each of these choices, N = 460.) in 57.3% of the instances the system's choices were identical to the experts 76.7% were identical or acceptable alternatives, and in 81.8% of the cases, its results were acceptable to the majority of the experts. In addition, the experts were asked to indicate organisms MYCIN may have overlooked. The experts indicated an average of only 0.36 organisms overlooked by the system. In no case did a majority of experts feel any single organism had been overlooked, suggesting that this figure of 0.36 may be a result of intra-expert variability.

The average number of organisms that MYCIN selected as requiring therapy was 3.83 (range 3-5). The average number of organisms that the experts selected was 6.09 (range 2-9). The experts did not always agree among themselves, but when there existed a consensus as to what organisms required therapy, MYCIN also covered for all organisms that a majority of experts felt must be covered for.

<u>C.</u> <u>Therapy</u> <u>Selection</u>. Each expert was asked to select therapy for the organisms which he felt were likely to be present before looking at MYCIN's therapy recommendation. He was then asked to judge MYCIN's choice of therapy for that patient. Since MYCIN was selecting therapy for the organisms which it felt were present (which may have differed from those chosen by the expert), this provides a fundamental comparison of performance - it compares therapy selection performance of the two when they are faced with the same clinical situation.

In those patients judged as requiring antibiotics, MYCIN selected an average of 1.5 drugs per patient (range 1-2) while the experts selected an average of 1.6 drugs per patient (range one - three). In the cases requiring treatment, the majority of the experts were willing to accept MYCIN's recommendation as a satisfactory alternative 73% (8/11) of the time. In 9% (1/11), the majority of the experts rated MYCIN's therapy as unsatisfactory. In 18% (2/11), the experts were equally divided as to whether MYCIN's therapy selection was satisfactory or unsatisfactory.

Table 1 shows that 15.5% of the time MYCIN's recommendation was identical to that of an expert. The experts were also asked to judge whether the system's therapy was an acceptable alternative (or identical to their own) producing the 73.6% figure in the second column.

The experts agreed among themselves (39% of the time) more than they agreed with MYCIN (28%). The results were analyzed to see if there were significant differences in ratings given by Stanford experts versus non-Stanford experts. In 9/11 cases the Stanford experts' ratings of therapy selection were similar to that of the non-Stanford experts. However, in the two remaining cases, four of five Stanford experts rated MYCIN's therapy selection as satisfactory, while four of five non-Stanford experts rated MYCIN's therapy as unsatisfactory.

<u>D. Overall Performance</u>. At the end of each evaluation form, the expert was asked to rate the system's overall performance as either excellent, good, fair, or poor. The first two columns of table 2 indicate that 17% of these evaluations were "excellent", and almost 60% were either "excellent" or "good" (only 13% were "poor"). In 60% of the cases (9 out of 15), a majority of the experts felt that MYCIN's overall performance was either "excellent" or "good".

<u>E. Present Utility and Future Potential</u>. Finally, after completing the entire set of 15 patients, each expert was asked to rate MYCIN's present utility and future potential as a clinical tool and as an educational tool, rating it as having "considerable", "some", or "no" potential.

All of the experts said that MYCIN has "considerable" or "some" potential as both a clinical and an educational tool in the future. (One of ten said "considerable" potential as a clinical tool, seven of ten said "considerable" potential as an educational tool.) At present, however, two of the ten experts said that MYCIN had no utility as a clinical tool and only one of ten said that it has "considerable" utility. Also, only one said that it has considerable present utility as an educational tool as well.

4 <u>CONCLUSIONS</u>

Experience to date suggests that codifying individual decision rules separately offers a large number of advantages, including flexibility and comprehensibility. It can provide the basis for a formalism capable of functioning in domains where little statistical data is available, or where information is uncertain or incomplete, and can thus offer a useful extension to existing techniques. Evaluation of MYCIN by outside experts also suggests that reasoning programs based on this methodology are capable of high performance.

Table 1. Summary of Ten Experts' Responses to MYCIN's Performance on 15 Cases.

	% of instances MYCIN's first choice was identical to an expert's first choice	% of instances MYCIN's first choice was identical to, or judged an acceptable alternative to an an expert's first choice	% of cases MYCIN's first choice was identical to, or judged an acceptable alternative by a majority of experts
ORGANISM SIGNIFICANCE	92.5% N = 40	NA	100% N = 4
ORGANISM IDENTITY	57.3% N = 460	76.7% N = 460	81.3% N = 11
THERAPY SELECTION	15.5% N = 110	73.6% N = 110	73% N = 11

Table 2. Summary of Ten Experts' Overall Ratings of MYCIN.

	% of times MYCIN was rated excellent by one expert	% of times MYCIN was rated excellent, good, or fair by one expert	% of times MYCIN was rated excellent, or good by a majority of experts
OVERALL PERFORMANCE	17.0% N = 150	87.2% N = 150	60.0% N = 15

REFERENCES

[1] Reiman H H, D'ambola J, The use and cost of antimicrobials in hospitals, Arch Environ Health, 13:631-636 (1966).

[2] Kunin C M, et.al., Use of antibiotics: a brief exposition of the problem and some tentative solutions, Anns Int Med, 79:555-560 (1973).

[3] Sheckler W E, Bennett J V, Antibiotic usgae in seven community hospitals, J Amer Med Assoc, 213:264-267 (1970).

[4] Roberts A W, Visconti J A, The rational and irrational use of systemic antimicrobial drugs, Amer J Hosp Pharm, 29:828-834 (1972).

[5] Simmons H E, Stolley P D, This is medical progress? Trends and consequences of antibiotic use in the United States, J Amer Med Assoc, 227:1023-1026 (1974).

[6] Kagan B M, Fanin S L, Bardie F, Spotlight on antimicrobial agents, JAMA, 226:306-310 (1973).

[7] Meyer A U, Weissman W K, Computer analysis of the clinical neurological exam, Computers and Biomedical Research, 3:111-117, (1973).

[8] Warner H R, Toronto A F, Veasy L G, Experience with Bayes's Theorem for computer diagnosis of congenital heart disease, Anns NY Acad Sci, 115:558-567, (1964).

[9] Gorry G A, Barnett G O, Experience with a model of sequential diagnosis, Computers and Biomedical Research, 1:490-507, (1968).

[10] Edwards W, N = 1, diagnosis in unique cases, Computer Diagnosis and Diagnostic Methods, (Jacquez, ed.), pp 139-151, C C Thomas, Springfield, Illinois, (1972).

[11] Shortliffe, E.H., Computer-Based Medical Consultations: MYCIN, New York: Elsevier, 1976.

[12] Davis R, Applications of meta level knowledge to the construction, maintenance, and use of large knowledge bases, A I Memo 283, Computer Science Department, Stanford University, July 1976.